**ONLINE | JNTU KUKATPALLY | AMEERPET**

**­­­­­­­­­­­­­­­­­­­­**

**Data Science Course Content**

**By VENKAT**

**@** [**learningpath17@gmail.com**](mailto:learningpath17@gmail.com)

**@+91 9966017501**

**Trainer Name : Venkat**

**Total Exp : 14+ years**

**Relevant Exp : 7+ years**

**No of Days Training : 60 days (90 Hours)**

**Sample Project : Yes**

**Fee :**

**Introduction to DataScience / Analytics**

* **Why does companies need Data scientist/Analyst?**
* **Data Analytics: OLAP vs Data Mining**
* **What is Data science? Why Data science?**
* **What is Data driven product engineering**
* **How to become a Data scientist?**
* **Skillset of Data scientist?**
* **Career Opportunities & Hiring companies.**

**Business Problems with Data Science and Data types**

* **Predictive Analytics Problems: Classification, Regression, Recommenders. (Supervised techniques)**
* **Descriptive Analytics Problems: Frequent Pattern Mining, Clustering, Outlier Detection. (Un Supervised techniques)**
* **Prescriptive Analytics problems: Predictive and Descriptive Problems.**
* **Types of Data:**
* **Structured data**
* **Unstructured data**
* **Semi structured data**
* **Time Series data**
* **Business Verticals: Retail, Banking, Financial, Auto mobile, Social, Web, Medical, Scientific, Logistics, Real Estate etc.**

**Required Technologies/Tools for Data Science**

* **Data science Life Cycle for Analysis**
* **Required technologies for each phase of Data Science life cycle.**
* **Single Machine Analytic Platforms: R, Python, SAS, etc...**
* **Distributed Analytical Platforms: Hadoop, Spark, H20**

**Mastering in Python language**

* **Python introduction and Installation**
* **IDEs set up**
* **Python basic topics**
* **Variables (Number, Strings, List, Tuple, Dictionary and Sets)**
* **Decision making**
* **Loops**
* **Functions, etc.**
* **Python advance topic**
* **Classes and OOPs Concepts**
* **Modules & packages**
* **File Handling**
* **Database handling, etc.**
* **Exception handling**
* **Python advanced features**
* **Required Packages for Data science in Python**
* **NumPy**
* **Pandas**
* **Matplotlib**
* **Sklearn, etc.**

**Statistics and Mathematics for Data scientist/Analyst.**

**Statistics**

* **Descriptive stats for single variable**
* **Mean, Median, Mode, Quantiles, Percentiles**
* **Standard Deviation, Variance**
* **MAD, IQR**
* **Descriptive stats for two variables**
* **Covariance**
* **Correlation**
* **Chi-squared Analysis**
* **Hypothesis Testing**
* **Inferential Statistics**

**Linear Algebra**

* **Ideas that need Linear Algebra**
* **Vector Algebra**
* **Ideas that map to vectors**
* **Understanding vector operations**
* **Matrix Algebra**
* **Ideas that map to matrices**
* **Understanding matrix operations**
* **Understanding eigen-values and eigen-vectors**
* **Concepts of basis**
* **Understanding factorization & Types**
* **Spectral factorization**
* **Eigen factorization**
* **SVD factorization**

**Probability**

* **Basic Probability**
* **Conditional Probability**
* **Bayes Rule/Reasoning**
* **Mapping Random process to Random variable**
* **Properties of Random variables**
* **Probability Expectation**
* **Entropy and cross-entropy**
* **Estimating probability of Random variable**
* **Understanding standard random processes**
* **Understanding on Probability Distributions**

**Calculus for Data Scientist**

* **Rate of change**
* **Concept of limit**
* **Concept of derivative**
* **Partial derivatives & gradient**
* **Significance of gradient**
* **Concept of integration, etc.**

**Data Visualization**

* **Tabular form**
* **Using statistical methods – mean, medium, mode, range, frequency, multi-dimensional tables, etc.**
* **Graphical form**
* **Bar graphs**
* **Histograms graphs**
* **Pie graphs**
* **Area graphs**
* **Density graphs**
* **Scatter graphs**
* **Line graphs**
* **Whisker graphs**
* **Correlation graphs**
* **Facet plots, etc.**

**Overview of Machine Learning**

* **What is Machine Learning?**
* **ML – Software Development Life Cycle**
* **ML-SDLC Phases**
* **Data Collection**
* **Data Preparation**
* **Feature Engineering**
* **Model Building,**
* **Model Evaluation**
* **Model Deployment**
* **Model Maintenance**
* **Type of Machine Learning Algorithms**
* **Supervised**
* **Unsupervised**
* **Semi-supervised**
* **Reinforcement Algorithms**

**Data Collection Techniques**

* **Collecting data from Excel/csv/txt files**
* **Collecting data from databases**
* **Collecting data from services**
* **Collecting data via scraping (from Web)**

**Data Preparation Techniques**

* **Structured Data Preparation**
* **Handling Missing Data**
* **Data Type Conversion**
* **Category to Numeric Conversion**
* **Numeric to Category Conversion**
* **Data Normalization:0-1, Z-Score**
* **Handling Skew Data: Box-Cox Idea**
* **Text Data Preparation/preprocessing**
  + **Noise removal (Stop word removal, URLs, punctuations, etc.)**
  + **Lexicon Normalization (Stemming & Lemmatization)**
  + **Object Standardization (Convert acronyms to dictionary words, grammar check, spell check etc.)**

**EDA (Numerical & Graphical) and Feature Engineering**

* **Exploring Individual Features**
* **Exploring Bi-Feature Relationships**
* **Exploring Multi-Feature Relationships**
* **Create new Features.**
* **Feature/Dimension Reduction: PCA**
* **Intuition behind PCA**
* **Covariance & Correlation**
* **Relating PCA to Covariance/Correlation**
* **Intuition to math**
* **Applications of PCA: Dimensionality Reduction, Image Compression**

**Model Building (ML Algorithm Building)**

* **Mathematical understanding for each model**
* **Limitations for each model**
* **Tuning for each model**
* **Model scope for type of problem: Classification, Regression Recommenders and Association.**
* **Pros and Cons of each model**

**Supervised learning Models**

* **Decision trees**
* **Probability learning (Naive Bayes)**
* **KNN Learning**
* **Linear regression**
* **Non-Linear regression**
* **Logistic regression**
* **Support Vector Machines(SVM)**
* **Ensemble Models**
* **Bagging**
* **Bagged trees**
* **Random Forest**
* **Extreme tress**
* **Boosting**
* **Ada boosting**
* **Gradient Boosting**
* **Extreme Gradient Boosting**
* **Voting (by Stacking aggregation)**
* **Soft voting**
* **Hard voting**
* **Stacking**
* **Neural Network Model**

**Unsupervised Learning**

* **Clustering Models**
* **K-Mean Model**
* **K-Medoid Model**
* **K-Centers Model**
* **Hierarchical Model**
* **Dimension reduction: Principal component analysis (PCA)**

**Association Models.**

* **Apriori Model (Association model)**

**Time Series Models**

* **Holt-Linear**
* **Holt-Winters (Extension for Holts Linear for Trend and Seasonality)**
* **ARIMA Model**
* **SARIMAX**

**Model Evaluation & Implementation Techniques**

* **Repeated holdouts (R.H)**
* **K-fold Cross-validation**
* **Bootstrap**
* **Metrix based evaluation.**
* **Model Implementation Procedures**

**Overview of Distributed/BIGDATA Analytics**

* **Big data Analytics Overview**
* **Platforms for Distributed Analytics: Hadoop, Spark, H20**
* **Hoop and Spark Overview & Comparison**

**Highlights**

* **Introduction for AI Algorithms**
* **Daily Practice with topic level exercises**
* **Mini Project**
* **Provide material for Python & ML (PPTs and Topic wise codes)**
* **Sessions on Interview questions.**
* **Sessions on resume preparatio****n.**
* **Mock interviews and job assistance.**